Test Instructions

- This test starts at the 2:35 PM. You have until 4:55 PM to finish.
- This is an open-book, open-note test. However, do not contact any "live" non-staff person via electronic or other means while you take this test.
- Please create your answers in electronic format - just as you have for lab assignments - and save your work, as usual, in your personal locker (i.e., the I:\ drive) or in a USB drive. You can edit your answers directly into your copy of this test if you have an HTML editor. For your convenience we have also provided MS-Word and PDF versions of the test in case you find these formats easier to use.
- Even where we don't require an explanation, you may want to say a few words about what you did so that you can get partial credit if some of your numerical answers and names are incorrect but your method has merit.
- Turn in your test (both the text and the PDF exports of your maps) by uploading your answer documents to Stellar (just as you have done for homework). If you wish, you may open the answer sheet in MS-Word and include your PDF images within your answer sheet as a single document.
- Remember to turn in your test before leaving the room. Please confirm with an instructor that your exam has been received. We strongly recommend that you retain the file containing your answers until we return the graded exams to you.
- Make sure to include your name and Athena user ID near the beginning of each document that you turn in.
- Finally, don't spend all your time on one or two questions. Start by looking over all the questions to get a general sense of what, and how much, is being asked, and the grade points associated with each question--then start work on your answers. Move on to the next question if you've spent more than 10 minutes on any short answer question; mapping questions will likely take longer.

Good luck!

Datasets for the Test

For this test, we will use US county-level data from the November 2016, US presidential election (between Donald Trump and Hillary Clinton). The relevant data for the test is bundled together in a folder called, test18_data, in the class locker. (Yes, these same data were used in last year's test, but we will focus this year on Texas results.) Copy the entire (read-only) folder from Z:\afs\athena.mit.edu\course\11\11.188\test18_data to a writeable space on your local drive (e.g., C:\TEMP).

Look in your copy of the folder for an ArcMap document called 11.188_test18_start.mxd
that has already included some of the shapefiles and tables that you will need. This ArcMap Document utilizes one new Metropolitan Statistical Area shapefile from the US Census plus several shapefiles that were obtained from the MIT Library Geodata Repository and then projected to a North_America_Albers_Equals_Area_Conic projection (NAD 1983) projection (in meters). (The standard EPSG code for this coordinate system is 102008 so I have included it in the shapefile names, but you will not need to use this fact.) For each of these shapefile, only the contiguous 48 United States were saved (the "lower 48").

<table>
<thead>
<tr>
<th>Filename</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>us_a1city_2000_lower48_epsg102008.shp</td>
<td>Point feature shapefile showing the location of major US cities (within the lower 48 states) having 2000 US Census population of at least 50,000.</td>
</tr>
<tr>
<td>us_f7states_2006_lower48_epsg102008.shp</td>
<td>Polygon feature shapefile showing state boundaries for the lower 48 states.</td>
</tr>
<tr>
<td>us_e25msa_2010_lower48_epsg102008.shp</td>
<td>Polygon feature shapefile from US Census Bureau of the US 2010 Census boundaries of metropolitan statistical areas (MSAs). The original shapefile has been limited to the lower 48 states and projected to the EPSG=102008 coordinate system.</td>
</tr>
<tr>
<td>us_f7counties_1996_lower48_epsg102008.shp</td>
<td>Polygon feature shapefile from US Census Bureau of the 1996 County boundaries within the lower 48 states together with the state, county name, Federal Information Processing Standard code (FIPS), and estimated 1990 and 1996 population, projected to EOSG=102008.</td>
</tr>
<tr>
<td>USinterstates.shp</td>
<td>Line feature shapefile of the US interstate highways. Obtained from ESRI (the ArcGIS vendor) and stored in geographic coordinates (NAD83). Not required for any questions but available, if desired, for context.</td>
</tr>
<tr>
<td>USboundary.shp</td>
<td>Line feature shapefile of US borders converted to EPSG=102008 after extraction from World country border shapefile downloaded from Natural Earth (<a href="http://www.naturalearthdata.com/downloads/110m-cultural-vectors">http://www.naturalearthdata.com/downloads/110m-cultural-vectors</a>).</td>
</tr>
<tr>
<td>2016_US_County_Level_Presidential_Results.csv</td>
<td>A comma-separate-value (CSV) text file containing the 2016 US Presidential Election results by County showing the votes for Donald Trump and Hillary Clinton, the total votes, the percentage of votes for these two candidates, the difference in votes, and the percentage point difference in their votes.</td>
</tr>
<tr>
<td>11.188_election16.mdb</td>
<td>MS-Access database containing two key tables: (1) election16_county with county-level presidential election results plus a few 5-year 2015 ACS census variables, and (2) election16_county_data_dictionary explaining the meaning of each field in election16_county.</td>
</tr>
</tbody>
</table>

Within the ArcMap document, 11.188_test18_start.mxd, the several state, county, MSA, and city shapefiles have already been added. The counties are shown thematically using quantile classification broken into 10 categories based on the 1996 population of each county. The county boundaries are omitted in order to avoid map clutter and make the map more readable. Data are shown for the 3111 counties in the lower 48 states.

In addition to the shapefiles, spreadsheet, CSV file,, and ArcMap document, the test18_data
folder also contains an MS-Access database, **11.188_election16.mdb**, with the **ELECTION16_COUNTY** table of county-level presidential election results along with selected census data extracted from the 5-year 2015 ACS census. For your information, the MS-Access database also contains a data-dictionary, **election16_county_meta**, explaining the columns in the election16_county table.

The 2016 election dataset was obtained from [https://github.com/tonmcg/County_Level_Election_Results_12-16](https://github.com/tonmcg/County_Level_Election_Results_12-16). Portions of the CSV file with election data were then imported into MS-Access and joined with ACS census data to produce **ELECTION16_COUNTY**. The 5-year 2015 ACS census data included with the election data come from the S1701 table on Poverty Status and the S2301 table on Employment and have been downloaded from the ACS website: [https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk](https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk). In addition, we have added population counts for years 2000 and 2010 to the MSA shapefile, **us_e25msa_2010_lower48_epsg102008.shp**. These population data were obtained from an Excel spreadsheet downloaded from: [https://www.census.gov/population/www/cen2010/cph-t/cph-t-7.html](https://www.census.gov/population/www/cen2010/cph-t/cph-t-7.html). [In order to join the population data for Los Angeles with the most similar shapefile polygon, we had to change the CBSA_code for Los Angeles-Long_Beach-Santa_Ana from 31100 to 31080 (Los_Angeles-Long_Beach-Anaheim). This editing issue is not relevant for the test.]

---

**Part I: Short Queries (36 Points)**

**Question I-1 (24 points total, 4 points each part) - Queries of election16_county**

- How many counties are there in Texas? ______________
- How many votes did Trump receive in Texas? _____________. How many votes did Hilary receive in Texas? ____________
- What was the largest (absolute) vote difference between Trump and Clinton among Texas counties? diff = ____________, fips = _____________.
- Which candidate received the most votes in that 'big difference' Texas county: Candidate: _____________. Votes: ____________
- What is the smallest total number of votes received by Hillary Clinton in any Texas county? ____________ fips = _____________. Total number of votes in that county? ____________
- How many of the Texas counties favored Clinton over Trump? ____________

**Question I-2 (12 points total, 3 points each) - Queries about metropolitan statistical areas (MSA) in us_e25msa_2010_lower48_epsg102008.shp**

Note that some MSAs cross State boundaries and the MSA and State boundaries come from different sources and do not exactly line up even when the share a common boundary.

- How many MSAs in the entire US shapefile are completely within Texas? ____________
- How many MSAs in the entire US shapefile have their centroid within Texas? ____________
- How many MSAs in the entire US shapefile have a reported 2010 population (pop2010) greater than or equal to 250000: ____________?
- How many of the centroid-in-Texas MSAs also have pop2010 >= 25000 ? ____________
Part II: Mapping Concepts (8 Points)

Question II-1 (8 points total, 4 points each)

When you open the map document, the county map is displayed using a North_America_Albers_Equal_Area_Conic projection (NAD 1983) rather than in geographic latitude-longitude coordinates. As we know, local planning agencies typically use projected coordinate systems. For example, the MassGIS mapfiles that we have used are generally saved in the Massachusetts (mainland) State Plane coordinate system.

Part II-1a (4 points): Explain briefly why national and world geographic datasets are usually distributed in geographic coordinates (e.g., latitude/longitude data in decimal degrees using WGS84), but local agencies such as MassGIS and Boston or Cambridge prefer to store and display geographic datasets in State Plane coordinates. Provide at least one reason for each case.

Part II-1b (4 points): Suppose we were to view the same US County map in geographic coordinates. Compared with the orginal view (using EPSG 102008) would the lat/lon map looked stretched in the east/west (horizontal) direction or the north/south vertical direction? Explain briefly why this is the case. [Note: You are welcome to change the display coordinate system of your ArcMap window, but there is no need and no map to submit for this question. Just provide a brief answer and explanation. If you do change your display coordinates, we suggest that you do so in a different data frame since the maps we ask you to submit should be in EPSG=102008, which (in ArcMap) is North_America_Albers_Equal_Area_Conic projection (NAD 1983) in meters.]

Part III: Election Map (30 Points)

Let's develop a choropleth map showing the percentage of votes (per_gop) for Donald Trump in each county. Add the election16_county table to ArcMap, join it to the shapefile of 'lower-48' counties and design your map. Since we will only ask questions about Texas counties, you may want to bring the data for only those counties into ArcMap. Note that election16_county lists the county FIPS code as text (field=fips_txt) as well as an integer (field=combined_fips). We added the fips_txt field for your convenience. It is not in the original CSV file. Be careful to select the appropriate data type when you join the table to your shapefile.

In addition, please notice that election16_county has many columns, most of which are not needed for the test. We have made these available in case you want to explore this data beyond the test. This may be interesting for some of you, depending on your research interests.

Question III-1 (15 points total)

MAP #1: Prepare and submit an ArcMap layout of the counties in Texas states and shade the counties based on the percentage of votes that were for Donald Trump. Be sure to:

- Zoom in to a little beyond the Texas boundary, show the other State boundaries, and include the usual scale bar, source, legend, north arrow, and an appropriate title,
- Leave the map in the original (EPSG=102008) projection,
• Indicate your choice of classification method for your thematic map of 'per_gop',
• Overlay a point map of major cities (in the us_alcity_2000_lower48_epsg102008 shapefile) with the size of the city 'point' proportional to the population of the city. Only include cities that have POP >= 100000,
• Include only those MSAs that you determined earlier to have their centroid-in-Texas MSAs and also have MSA pop2010 >= 250000. To facilitate readability, make them 50% transparent and fill them with some type of 'hash' pattern,
• Adjust color ramps, choice of background and foreground colors, outlines, etc. to improve the readability of the map.

NOTE: This question is worth 15 points and provides an opportunity to demonstrate the cartographic skill that you have developed to develop an informative, readable map.

**Question III-2 (6 points total)**

**Part III-2a (3 points):** Briefly discuss your choice of classification scheme and number of categories for your election results map for Texas.

**Part III-2b (3 points):** Briefly discuss any spatial pattern that you observe in your map regarding Trump's percentage of votes among the Texas counties and MSAs.

**Question III-3 (9 points total)**

**Part III-3a (2 points):** What is the average (mean) value of the percentage of votes for Trump (per_gop) among all the Texas counties for which you have the voting results joined to your map:

___________?

**Part III-3b (3 points):** Select those counties whose centroids are within the 'centroid-in-Texas MSAs' that also have pop2010 >= 250000 - that is, one of the collections of MSAs that you examined earlier. How many counties are in your 'big metro' county selection? ________  What is the total number of votes for Trump _________ and for Clinton _________ within these 'big metro' counties?

**Part III-3c (4 points):** Using the county-level ACS census data joined to your county shapefile, determine the average (mean) percentage of persons below the poverty level (pct_below_pov) for the 'big metro' Texas counties _________ and for all the other Texas counties _________.

---

**Part IV: Proximity to the Mexican Border (26 Points)**

Next, let's examine the presidential voting pattern for counties near the US-Mexico border as shown in the US_boundary_epsg102008 shapefile.

**Question IV-1 (6 points total)**

Create a buffer of 50 km radius around the US-Mexico border. Select all the counties that intersect your buffer. (That is, the counties that have any or all of the county area within your border buffer.)

**Part IV-1a (4 points):** How many counties fall within your buffered large cities: __________?
Part IV-1b (2 points): Among those counties that overlap the large city buffer, what is the total number of votes for each candidate (i.e., the sum of votes_gop, and the sum of votes_dem): Trump votes: ____________, and Clinton votes: ______________?

Question IV-2 (7 points total)

Using the county-level ACS census data joined to your county shapefile, lets estimate the fraction of persons in each county who are categorized as Hispanic or Latino (of any race). Use 'pop_16plus' for the total number of persons aged 16 and older, and use 'pop_16p_hispanic' for the number of persons aged 16 and older who are of Hispanic or Latino origin (of any race).

Part IV-2a (4 points): What is the total number of persons aged 16 and older (pop_16plus) in those Texas counties within your 50 km buffer of the Mexican border ____________ and in those Texas counties beyond the 50 km buffer ___________ ?

Part IV-2b (3 points): Compute for every county in Texas a new variable, pct_hispanic, that is the percentage of persons aged 16 and older who are Hispanic or Latino (of any race). For the County of El Passo, FIPS=28141, in the Western most tip of Texas along the Mexican border, what is the pop_16plus __________, pop_16p_hispanic __________, and pct_hispanic __________?

Question IV-3 (13 points)

Create a scattergram chart that plots for every Texas county the pct_hispanic vs. pct_gop.

MAP #2 (10 points): Prepare and submit a second ArcMap layout of the Texas counties and again shade the counties based on the percentage of votes that were for Donald Trump. This time, be sure to:

- Include the outline of the States, and the usual scale bar, source, legend, north arrow, and an appropriate title,
- You may use many of the same layers as in the first Map, but add the following:
  - Overlay your 50 km buffer of the Mexico border with enough transparency to allow the chloropleth map of per_gop to be visible underneath.
  - Highlight those Texas counties that overlap any part of the border buffer.
  - Include in your layout the scattergram chart of Texas counties showing the percentage vote for Trump (pct_gop) vs. pct_hispanic, the percent of persons aged 16 and older who are Hispanic or Latino origin (of any race).
  - Adjust the symbols in your scatterplot chart so that those counties within the 50 km buffer are one color and all the other counties are shown in another color.

Part IV-3b (3 points): Briefly discuss any spatial pattern that you observe in your chart of pct_gop vs. pct_hispanic.

That's all for the test, but feel free to keep the election data and explore some of the election patterns further when you have time. We do not have time on the test to examine further the relationship between voting, demographics, and health exchange participation and subsidies.

Please note:
Before creating the PDFs, be sure to include your name and Athena ID somewhere on the text file and the map layouts.

You should test your PDF files with Adobe Acrobat Reader before submitting them to Stellar in order to be sure they are readable.

You should keep a copy of your text output and PDFs file in your Athena locker until your test is graded.

You should confirm with an instructor that your test files were received.

Last modified 11 March 20018. (jf)

Back to the 11.188 Home Page.
Back to the CRON Home Page.